



Research Report  
Statistical Research Unit  
Göteborg University  
Sweden

---

**Statistical measures for  
evaluation of methods  
for syndromic surveillance**

**Marianne Frisé**

**Research Report 2003:11  
ISSN 0349-8034**

---

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-773 12 74	Nat: 031-773 10 00	<a href="http://www.stat.gu.se/stat">http://www.stat.gu.se/stat</a>
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

## Statistical Measures for Evaluation of Methods for Syndromic Surveillance

Marianne Frisé

Statistical Research Unit, Göteborg University, Sweden

### Abstract

#### Introduction

In syndromic surveillance there is a need for continual observation of one or more time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred. Statistical methods are necessary to separate important changes from stochastic variation. The statistical methods suitable for this differ from the standard hypothesis testing methods. Also the measures for evaluation differ.

#### Objectives

An overview of statistical optimality issues and statistical measures for evaluation of prospective surveillance will be given. Timeliness and the control of false alarms are important issues.

#### Methods

*Surveillance methods.* Some commonly used methods for surveillance are examined. *Optimality criteria.* Most of the commonly used methods are optimal in some respect. Different criteria of optimality are used in different subcultures of statistical surveillance. The shortcomings of some criteria of optimality are demonstrated. One criterion discussed is based on the average run length, ARL. This is the most commonly used optimality criterion. Another criterion is based on a utility function. From the perspective of optimal decisions costs are given to the different errors which can be made and a utility function is maximized. A third criterion discussed is that of

minimax. *Evaluation measures.* Several measures for false alarms, detection ability and predictive value will be discussed and illustrated.

## Conclusions

Evaluation of methods for syndromic surveillance in practice is very important. It is necessary to know the basic properties of a system before it is implemented. This involves many important aspects. The use of relevant statistical measures is one of them.

## 1. Statistical surveillance

For syndromic surveillance there are many important issues, such as data collection and data quality, to consider. Here, the focus will be on the statistical issues. Statistical methods are necessary to separate important changes from stochastic variation. The statistical methods suitable for this, differs from the standard hypothesis testing methods (1). Also the measures for evaluation differs (2).

Statistical surveillance deals with statistical methods to separate important changes from stochastic variation. See **Figure 1**.

Surveillance is characterized by: repeated measurements, repeated decisions, no fix hypothesis and that time is important. We would like to detect an outbreak early rather than late. Time is important for the construction of methods and also for the measures of evaluations which are relevant.

## 2. Evaluations

Evaluations are important. When monitoring is used in practice, knowledge about the properties of the method is important. If an unusual event occurs it is otherwise hard to know how serious you should take it.

In applied work a single optimality criterion is not always enough but evaluations of different properties might be necessary (3). The performance of a method for surveillance depends on the time  $\tau$  of the change. Alarm probabilities will in general not be the same for early changes as for late changes. Sometimes it is appropriate to express the measure of the performance as functions of time, as in (4).

First, we look at a standard situation of a sudden change in distribution at a certain change point time  $\tau$ . See **Figure 2**. The observation could be an age adjusted prevalence or some other derived statistic depending on the specific situation. Good properties are quick detection and few false alarms (3).

### 2.1 False alarms

The type I error is more complicated at surveillance than at hypothesis testing. This error depends on how long the surveillance runs. All reasonable methods for surveillance will have size 1 if the surveillance is run long enough. Special measures of the false alarm properties which are suitable for surveillance are suggested:

The Average Run Length at no change,  $ARL^0 = E(t_A | D)$ . A variant of this is the Median Run Length, MRL.

The false alarm probability  $P(t_A < \tau)$ . This is the probability that the alarm occurs before the change. In theoretical work, the standard procedure is to assume that  $\tau$  is geometrically distributed, implying a constant intensity of a change.

## 2.2 Delay of the alarm

Sensitivity depends on the time of the change and the length of time series. Thus there is not one unique sensitivity value at surveillance but other measures might be more useful.

The delay  $t_A - t$  should be as small as possible. The expected value is

$$ED = E_{\tau} E_{t_A}[\max(0, t_A - t) \mid \tau=t]$$

The most commonly used measure of the delay is the Average Run Length until detection of a change (that occurred at the same time as the inspection started).

$$ARL^1 = E[t_A \mid \tau=t]$$

Sometimes there is a limited time available for rescuing action. The Probability of Successful Detection measures the probability of detection with a delay time no longer than  $d$

$$PSD = P(t_A - \tau < d \mid t_A \geq \tau),$$

It is thus a function both of the time of the change and the length of the interval in which the detection is defined as successful.

## 2.3 Predictive value

The probability that there actually is a change at the alarm signal is

$$PV(t) = P(t_A \leq \tau \mid t_A = t)$$

Certain methods have a constant PV. Others have a low PV at early times but better later. In those cases, the early alarms will not motivate the same serious action as later alarms.

## 2.4 Computation

A free computer program that gives evaluation graphs is available from the author. It is self-instructing with extensive help-functions.

### **3. Optimality**

To choose the best method you have to consider what "best" means. Optimality plays an important role both in applied work and for theory (5). There are many papers which claim to give the optimal method of surveillance. However, the suggested optimality criteria differ in important aspects (2). You should choose method (and parameters in the method) which are optimal for the specific aim at hand. The requirements are different for short-term high-risk and long-time low-risk

#### ***3.1 ARL Optimality***

In the literature on quality control, optimality is often stated as minimal  $ARL^1$  for fixed  $ARL^0$ . It has been demonstrated (2) that useless methods are ARL optimal. Thus this optimality should only be used with care. The ARL can be used as a descriptive measure and give a rough impression but is dangerous as a formal optimality criterion.

#### ***3.2 Minimal expected delay***

The expected delay from a change to the detection is minimized for a fixed false alarm probability. This criterion also gives the minimum of a very general cost function.

#### ***3.4 Minimax Optimality***

The third criterion is the minimax of the expected delay after a change. It concerns the worst

value of the change point  $\tau$  and for the worst history of observations before  $\tau$ . This criterion is pessimistic since it is based on the worst possible circumstances. Much theoretical research is based on this criterion.

#### 4. Methods

We will not give the formulas for common methods but just report their respective optimality properties. This will indicate that different methods are good for different tasks

The full likelihood ratio method LR is optimal with respect to the minimal expected delay (5).

The EWMA method is approximately optimal with respect to the minimal expected delay for a certain value of the parameter of the method (2, 6).

The Shewhart method, which is similar to doing repeated significance tests, is optimal for a recent large change (5) but not to detect longterm smaller changes.

The CUSUM method is minimax optimal (7)

#### 5. Syndromic surveillance

When many symptoms are considered we have a situation of multivariate surveillance. If the incidence of the different symptoms change at the same time (or with a known time lag), then the multivariate situation is easily reduced to a univariate one (8). Also in other cases there are several ways to construct methods (2). The multivariate methods can also be evaluated as described above. However, optimality is always complicated in multi-dimensional cases.

## 6. Outbreak detection

At surveillance of the incidence, standard methods for the normal distribution are not suitable. However, there are methods available also for other distributions such as e.g. the Poisson distribution (1).

The situation can be multivariate because of many symptoms as discussed above. It can also be multivariate because of a spatial perspective.

At an outbreak the incidence typically increases gradually and then possibly declines. See **Figure 3**. The change is more complicated than the standard situation with a sudden shift from one level to another. It might be hard to model exactly the shape of the rise and the decline - or even to estimate the baseline accurately

A new robust method might be of interest. Frisé (9) suggested surveillance that is not based on any parametric model but only on monotonicity restrictions. The estimation procedure is the regression under order-restrictions (10). The surveillance method was described and evaluated by Andersson (11) and its use for outbreak detection is discussed by Andersson in this proceeding (12). The method is developed for cyclical processes and the aim is to detect a turn (peak or trough) as soon as possible. The method is based on the likelihood ratio.

The type of change detected is a rise in incidence at the outbreak. The method can also be used to detect the decline. This might be of interest e.g. to decide when an influenza epidemic is over and new cases with similar symptoms should give an alert.

## 7. Concluding remarks

Knowledge of the properties of a system for syndromic surveillance is very important both



for the choice of appropriate method and for the interpretation of an alarm.

This involves many aspects. One of them is the use of statistical measures which take care of the special time dependencies of a surveillance system.

Another important aspect of a system for syndromic surveillance is the robustness against miss-specification.

## References

1. Sonesson, C. and Bock, D. (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society Ser A*, **166**, 5-21.
2. Frisén, M. (2003) Statistical surveillance. Optimality and methods. *International Statistical Review*, **71**, 403-434..
3. Frisén, M. (1992): Evaluations of methods for statistical surveillance. *Statistics in Medicine*, **11**, 1489-1502.
4. Frisén, M. and Wessman, P. (1999). Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics, Simulations and Communications*, **28**, 597-622.
5. Frisén, M., and de Maré, J. (1991), "Optimal Surveillance," *Biometrika*, **78**, 271-280
6. Frisén, M. and Sonesson, C. (2003): Optimal surveillance by exponentially moving average methods. Submitted.
7. Moustakides, G. V. (1986), "Optimal Stopping Times for Detecting Changes in Distributions," *The Annals of Statistics*, **14**, 1379-1387.
8. Wessman, P. (1998) Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods*, **27**,

1143-1161.

9. Frisén, M. (1994) Statistical Surveillance of Business Cycles. 1994:3, Department of Statistics, Göteborg University.
10. Frisén, M. (1986) Unimodal regression. *The Statistician*, **35**, 479-485.
11. Andersson, E. (2002) Monitoring cyclical processes - a nonparametric approach. *Journal of Applied Statistics*, Vol. 29, pp. 973-990.
12. Andersson, E. (2003) A monitoring system for detecting starts and declines of influenza epidemics. *2003 National Syndromic Surveillance Conference*, New York, USA.

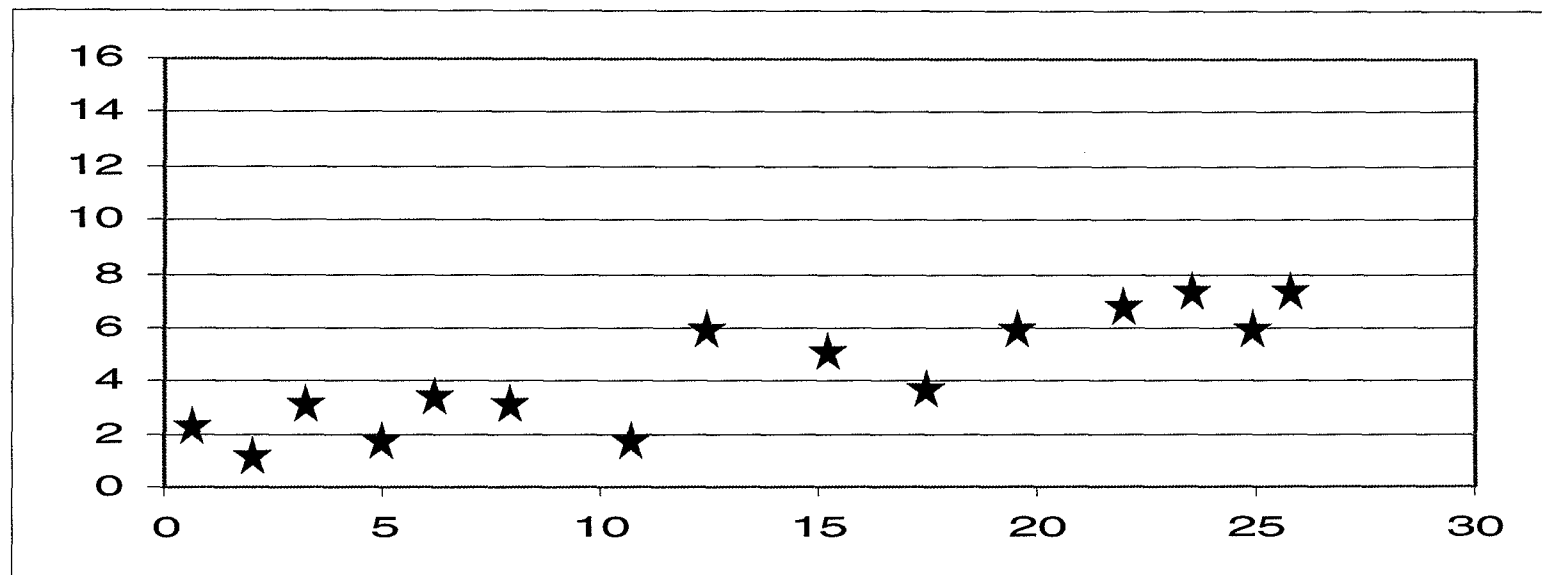


Figure 1. At which time do we have enough information to decide that the level has changed?

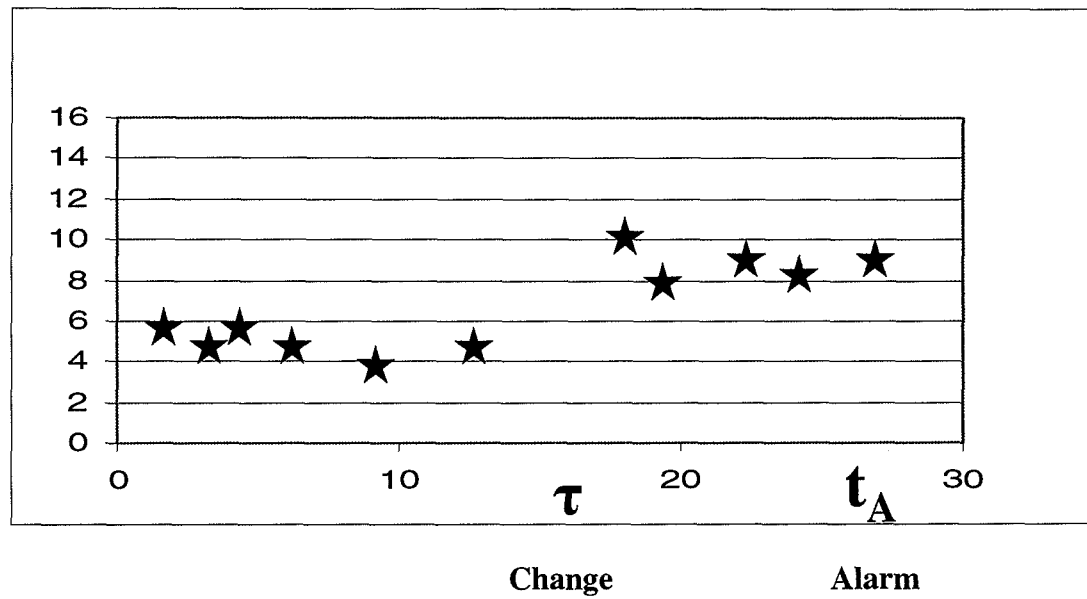


Figure 2. The First  $(\tau-1)$  observations  $x_{\tau-1} = x(1), \dots, x(\tau-1)$  have density  $f^D$ . The following observations have density  $f^C$

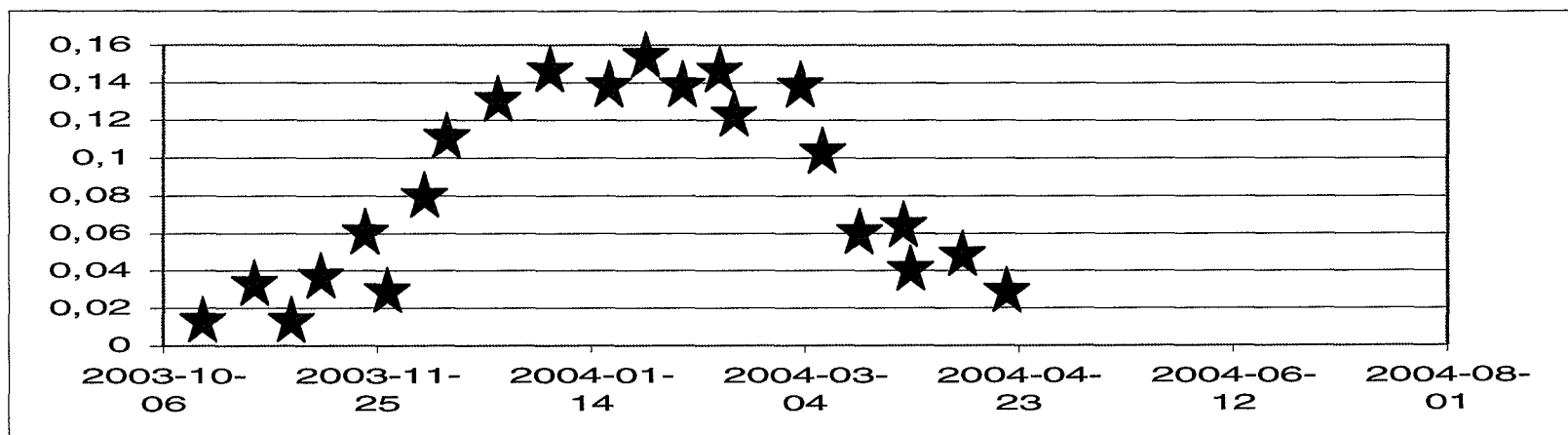


Figure 3. At an outbreak it is of interest to detect a rise and sometimes also the decline.

## Research Report

2002:7	Andersson, E., Bock, D. & Frisén, M.:	Some statistical aspects on methods for detection of turning points in business cycles.
2002:8	Andersson, E., Bock, D. & Frisén, M.:	Statistical surveillance of cyclical processes with application to turns in business cycles.
2002:9	Holgersson, T.:	Testing for non-normality in multivariate regression with nonspherical disturbances.
2003:1	Holgersson, T. & Shukur, G.:	Testing for multivariate heteroscedasticity.
2003:2	Holgersson, T.:	Testing for multivariate autocorrelation.
2003:3	Petzold, M. & Sonesson, C.:	Detection of intrauterine growth restriction.
2003:4	Bock, D.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2003:5	Holgersson, T. & Lindström, F.:	A comparison of conditioned versus unconditioned forecasts of the VAR(1) process.
2003:6	Bock, D.:	Early warnings for turns in business cycles and finance.
2003:7	Lindström, F.:	On prediction accuracy of the first order vector auto regressive process.
2003:8	Petzold, M. & Jonsson, R.:	Maximum Likelihood Ratio based small-sample tests for random coefficients in linear regression.
2003:9	Petzold, M.:	Preliminary testing in a class of simple non-linear mixed models to improve estimation accuracy.
2003:10	Frisén, M. & Gottlow, M.:	Graphical evaluation of statistical surveillance.